Assignment 3 – Query Evaluation

B561 - Fall 2011

Due: 10pm Nov. 17

Total score: 40 points + 10 bonus points

Consider a database with the following schema:

Student (sid, name, age, gender, dept, GPA)

Course (cid, name, description)

Enrollment(sid, cid, term, grade)

grade is an integer with values between 0-100.

We are to evaluation a SQL query:

```
Select S.name
From student S, enrollment E
Where S.sid = E.sid and S.dept = "CS" and E.grade>95
```

The following are what's known about the database and database instance:

1. buffer is not sufficient to hold all pages of the student table or the enrollment table. The size of the buffer is around 5-10% of the size of the three tables combined.
2. there are 50 departments in the university, and the number of students in the CS department is close to the average student number in a department.
3. less than 0.2% students have grade above 90.

Please construct the most efficient evaluation plan under the following circumstances (an index is non-clustered unless it is explicitly stated that it is clustered), and explain why you think the plan is efficient. In other words, for each circumstance, you need to present an evaluation plan tree, and write a few paragraphs defending your choice of data access methods and join algorithms. If you make any assumption besides what's explicitly stated in the question, please clearly list your assumption.

Case 1. Hash index on S.dept, Hash index on E.grade, no other index available. (20 points)

Case 2. Clustered B+ tree index on E.sid, B+tree on S.dept, no other index available. (20 points)

Case 3. B+tree index on S.dept, B+tree index on E.grade, no other index available. (10 bonus points)